Neural simulation of a system that learns representations of sensory experience

Author
Henry, Howard N.

Publication Date
1987

Peer reviewed

NEURAL SIMULATION OF A SYSTEM THAT LEARNS REPRESENTATIONS OF SENSORY EXPERIENCE

Howard N. Henry

Information and Computer Science Department
and
Center for the Neurobiology of Learning and Memory

University of California, Irvine, CA 92717

Technical Report #87-08

Abstract

The pyriform cortex forms stable representations of smells to allow their subsequent recognition. Clustering systems are shown to perform a similar function, so they provide a guide to understanding the operation of the pyriform. A neural model of a sample of pyriform cortex was built that adheres to most known biological constraints, including learning by long-term potentiation. Results of early simulations suggest some interesting properties. The effort has implications for the knowledge representations used in artificial intelligence work.

---

[1]The research reported in this paper would not have been possible without the expertise of Richard Granger, Gary Lynch, and Michel Baudry. This research was supported in part by the Office of Naval Research under grants N00014-84-K-0391 and N00014-85-K-0854, the National Science Foundation under grants IST-81-20685 and IST-85-12419, and by the Army Research Institute under grant MDA903-85-C-0324.

1 Computing Representations

How should an intelligent system organize its experience so that the stored knowledge can assist in future decisions? The simplest representations of a stimulus built by an intelligence permit it to *recognize* the re-occurrence of that stimulus. Recognition is difficult because each occurrence of a stimulus is slightly different, but a single representation must be formed for all these instances, past and future. This process is analogous to clustering of similar objects, in which we group individual dogs and cats, for example, into their respective categories. Thus it seems that the low-level knowledge representations used for recognition could be formed automatically by such clustering mechanisms.

Clustering is guided by similarities perceived internally, rather than by an external definition of category membership; this distinguishes it from "learning from examples" in the machine learning literature [Carbonell et al. 1983]. Many clustering techniques have been developed by machine learning researchers [Fisher 1987, Michalski & Stepp 1983], and by workers in the field of numerical taxonomy. While this paper is not concerned with these clustering methods *per se,* Fisher's [1987] COBWEB system implements an algorithm that optimizes "category utility" [Gluck & Corter 1985], a measure derived from information theory and the well-established psychological finding of "natural categories" or "basic level concepts" [Mervis 1981]. Category utility will provide a measure of the performance of the clustering systems to be studied, and Fisher's COBWEB can be used to find the optimum category utility for any set of data to be studied.

Clustering systems are typically applied to semantically-meaningful, symbolic inputs. This is inappropriate for our purpose, however, since it presupposes much of the answer to the question, "How should experience be organized?" It is quite possible that the "simplest" concepts revealed by introspection are actually elaborate compositions formed from the true, hidden building blocks.

So instead of working on clustering of *concepts,* this study will focus on what might be called "sensory clustering". Sensory systems have the advantage that they process observable external stimuli, and thus permit careful experimental controls on the inputs presented. In addition, mechanisms for sensory processing are relatively limited and localized in animals. But it is possible that sensory clustering does not conform to the same psychological constraints as conceptual clustering — e.g., the "basic level" constraint. Determining whether this is true will be one of the objectives of this effort.

Animal behavior and neurobiology will provide a specific physical system for testing this hypothesis. In particular, the study will address the recognition of smells by rats. The neurobiology of olfactory processing in rats is simpler and better-understood than most other sensory areas, so the detailed operations can be more completely specified for the study. And unlike early visual processing, where many of the operations are hard-wired from birth, the olfactory neurons learn from, and adjust to, the smell input sensations they have processed. The olfactory cortex can thus demonstrate how a system might organize itself according to its sensed experience.

2 The Overall Approach

Marr [1982] developed a three-level framework which is useful in studies such as this. The highest "computational theory" level describes the abstract defining constraints on a computation, such as Peano's axioms for addition. The middle "representation and algorithm" level specifies one of the many processes for carrying out a computation, such as appending a "0" to a binary number for the computation of doubling integers. The lowest "hardware implementation" level considers the multiple physical realizations (e.g., calculators, cash registers) for a process. Marr's point was that the computational theory needs to be considered in order to inform the interpretation of observations at the lower levels.

Marr's framework can be illustrated with the earlier conceptual clustering example. Category Utility Maximization provides a constraint that could be a computational theory for conceptual clustering. Fisher's COBWEB system uses one algorithm fitting this theory, and it is implemented in some specific LISP code.

This multi-level approach will be useful in this sensory clustering study. The rat's olfactory cortex provides an implementation of an unknown algorithm for smell processing. This algorithm may (or may not) conform to a computational theory of category-utility-maximization. The neural network can be simulated under varying conditions to find those conditions, if any, that yield an algorithm consistent with the computational theory. Then detailed neurophysiological measurements and behavioral experiments will verify whether the simulated conditions are met in the animal. The neural simulations permit a large number of alternatives to be ruled out without exhaustive experimentation.

This work is in its early stages. To this point, I have developed the neural simulation and observed some of the simpler properties of the representations produced by the simulation. This paper will focus on these results.

3 The System Being Modeled

The olfactory cortex performs the early processing of smells in mammals. It receives its principal input from the olfactory bulb over the "lateral olfactory tract" (L.O.T.); the bulb may serve to moderate the strength of the signals originating in the smell receptors in the nose. Output from the olfactory cortex travels over the perforant path to the hippocampus, where it is associated with inputs from other senses; the output also reaches several regions of the cerebral cortex. Four areas compose the mammalian olfactory cortex; this research concerns only one of those areas, the pyriform cortex, which is believed responsible for the recognition of smells [Lynch 1986].

The early processors of inputs from most other senses — vision, touch, limb position, and hearing — are topographically organized. Thus, relative positions in the input are preserved at successive stages of processing (Figure 1). Visual sensations, for example, are received on a "retinal map", and this map organization is maintained at several additional levels of computations. This organization is analogous to an image passing through a series of filters.

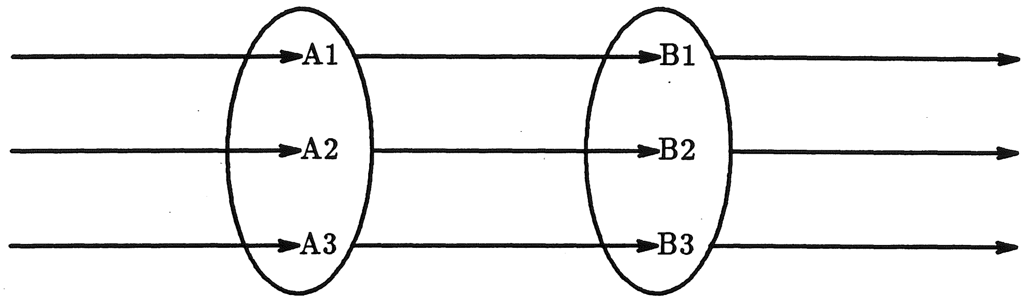

Figure 1: Topographic Organization of Processing

By contrast, processing of smells (and possibly tastes) are organized "combinatorially". Each L.O.T. input has a chance of contacting any pyriform cell [Lynch 1986]. Thus the smell inputs, with minimal topography at the bulb, are redistributed onto a new map in the pyriform, and new reorganizations are performed at each successive stage. The L.O.T. inputs effectively form an array of sparse, random connections with the pyriform cells.

The pyriform cells generate feedback onto themselves and their neighbors through another combinatorial array, similar to that for L.O.T. inputs. Figure 2 diagrams the organization described so far; the structure is very similar to the associative memory networks analyzed by Kohonen and his colleagues [1981], but the pyriform cells receive no "forcing stimuli" that impose an external organization. Superimposed on this system is feedforward inhibition from L.O.T. inputs onto pyriform cells, and feedback inhibition from pyriform cells onto themselves; this discussion will be clearer if details of the inhibitory system are omitted.

The most interesting property of this system is that it *learns*. At the cellular level, learning is manifested in changes in the strength of connections ("synapses") between the input and output. Synaptic weights in the two pyriform arrays described (as well as many other places in the mammalian brain) change according to the mechanism of "long-term potentiation" (LTP) [Lynch 1986]. LTP is consistent with the Hebb [1949] Rule, under which convergent coactivity of the inputs and output increases the strength of active synapses onto the output cell. Having modifiable connections enables the system to adapt itself to the input activity actually

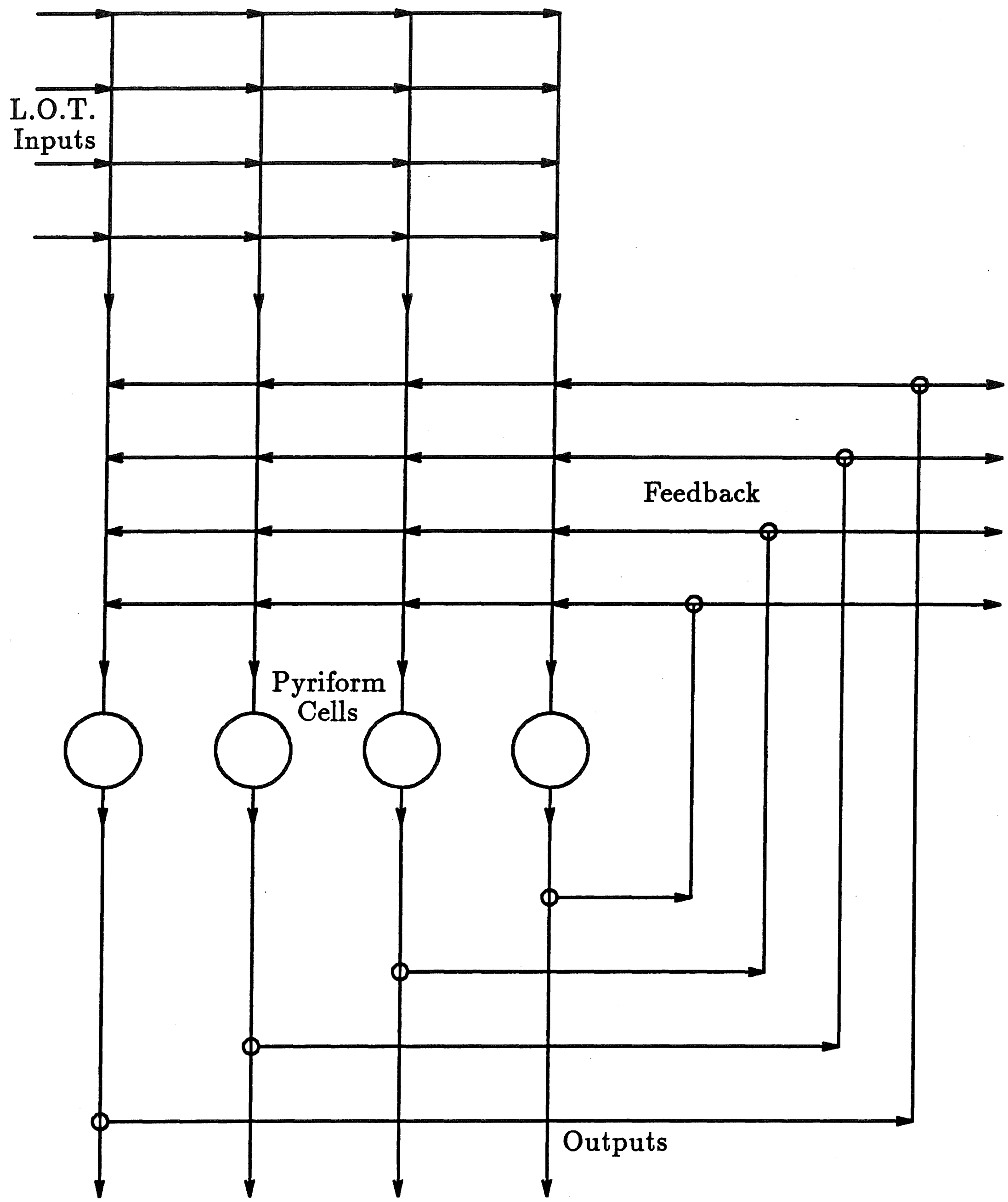

Figure 2: Pyriform Organization. The arrow intersections indicate a small chance of a connection, while the circled intersections show positive connections.

being experienced.

A simple example will illustrate how changing synaptic strengths cause the network to adjust its output representations in response to its experienced activity. In the beginning, a totally-unfamiliar stimulus will produce its independent built-in output. However, this experience can strengthen these connections so that similar later smells are more likely to evoke part of this output pattern than another output. Furthermore, this later experience can further strengthen the response to the overlap between the related stimuli. In this way, connection strengths will gradually move toward levels that reflect the relative co-occurrence of individual input features.

4 The Simulation Model

Three versions of computer simulations have been developed and used in this research:

1. An interactive C-language version is neurobiologist-friendly and flexible, but lacks some of the components of the pyriform cortex, such as the inhibitory system.

2. A batch C-language version, by contrast, is relatively complete but difficult to use and tailor to specific requirements.

3. The batch C-language version was also translated into FORTRAN, scaled up, and run on a Cray X-MP/48.

In the pyriform cortex, a million L.O.T. inputs connect sparsely with a million pyriform cells, and these million cells feed back through sparse connections with their million neighbors. To reduce memory and processing requirements for the simulation, a sample of only 100 L.O.T. inputs and pyriform cells was modeled. The strengths of the connections between the inputs and outputs are maintained in a 100×100 array, and another 100 by 100 array holds the feedback connection weights.

There are several levels of description possible in modeling a neural network, ranging down to the neurochemical level, where the ion flows across the cell membrane are calculated explicitly. This model is specified by the cell electrophysiology, such that electrical potentials are summed over the entire neuron, with all synapses being treated equally. Extensive biological details have been included to attain a realistic model at this level.

A minimal explanation of the operation of both the pyriform cortex and its model can assist in understanding this work (Refer to Figure 3):

1. Selected L.O.T. inputs become active.
2. The activation passes through connections, where they exist, to the receptors ("dendrites") of the pyriform output cells. The amount of activation flowing through depends on the learned strength of the connection.
3. The passed-through activation from all the connected inputs is combined and compared to that needed for learning; if this and the other LTP conditions are met, then the contributing synapses will be strengthened.
4. The combined activation is passed to the output cell, and compared to a threshold; if the threshold is exceeded, the output cell "fires", activating its feedback line and external outputs.
5. The feedback activation passes through other connections to the receptors of the output cells, like the L.O.T. input activity did.
6. The process cycles back to Step 3, and continues after the L.O.T. input vanishes until the feedback dissipates, usually due to growing inhibition not included in this explanation.
7. Later a new L.O.T. input appears, and the process returns to Step 1.

The simulation proceeds similarly, after it determines the connection arrays and the sequence of L.O.T. inputs. An outer loop processes each L.O.T. input, while an inner loop processes the feedback signals arising during that L.O.T. cycle. Within each loop, incoming activations are summed, firing is determined, inhibition is computed, and qualifying connections are strengthened. In addition to the raw output of pyriform cell activity,

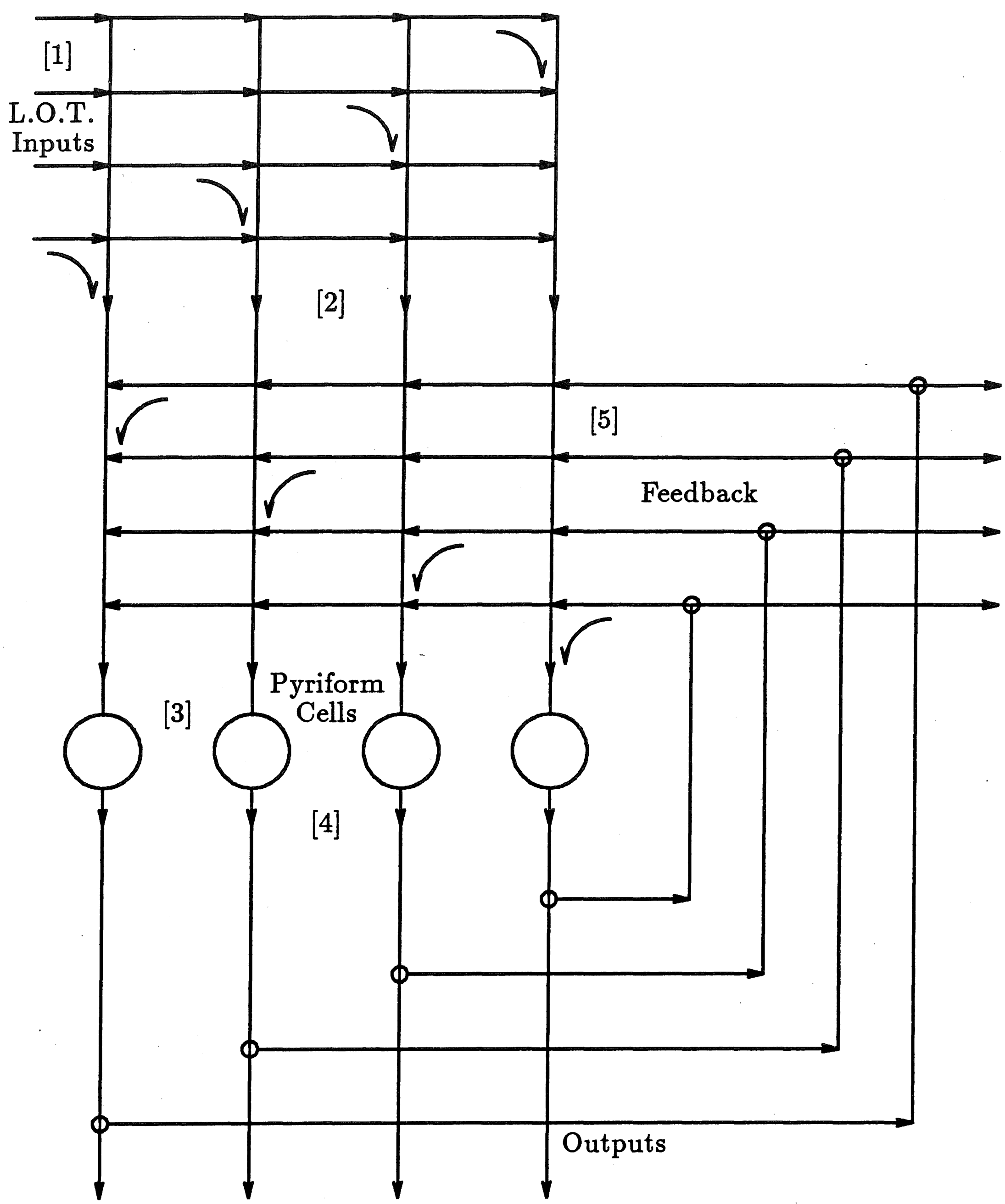

Figure 3: System Operation. The numbers in brackets show the general location for each step listed in the text.

Hamming distances are computed for each pair of inputs and each pair of outputs, and the paired input-output distances are graphed (Figure 4). The next section discusses some of the uses for this output.

5 What Can a Neural Simulation Show?

When a vector of L.O.T. activation is presented to the simulated system, some of the pyriform cells fire, producing an output vector of pyriform activity. This output vector behaves like a stable internal *representation* of the input, relative to the system's past experience. In essence, the system learns to recognize patterns that repeatedly co-occur in the environment. Characterizing the resulting representations is a major goal of this research.

One simple way of illustrating the functional relation between L.O.T. activity and the resulting representation is to answer the question: How does the output change as the input is varied (holding everything else constant)? Hamming distance may be used as a measure of the variation between two activation vectors. Plotting output differences against input differences could show linear, exponential, logarithmic, or S-curve relationships, among others (Figure 5).

The different functions have important implications on the performance of the representation. An S-curve relationship, for example, predicts that there is a threshold below which noisy inputs are mapped to the same result, and above which they are widely separated. A logarithmic curve, on the other hand, suggests that representations will be less and less differentiated as inputs vary, with no sudden threshold.

Preliminary simulations show a linear relationship between variations in input and variations in the corresponding internal representations (Figure 6), and this finding holds for a broad range of neurobiological parameters. This implies that the representations have the same relative similarities as their inputs, so the system is not performing a gross transformation on the inputs. Instead, it seems likely that the transformation has a subtler qualitative effect, not measurable with the Hamming distance metric. Two possible effects are *representational flexibility* and *categorization;* these will be discussed in the next sections.

```
v OUT    5    10   15   20   25   30   35   40   45
 21:                                1
 20:
 19:                               2  3
 18:                               13 63
 17:                                *  *3
 16:4                              8*42**
 15:4      1                       6*6***
 14:6    77                        2*****
 13:2    6*5                       ******
 12:*    ***                       **3***
 11:*    *2*                       ******
 10:*    ***                       ******
  9:*    ***                       ******
  8:*    ***                       *94*5*
  7:*    ***                       ***9*7
  6:*    ***                       *5*796
  5:*    ***                       55 759
  4:*    ***                       353126
  3:*     **                        4  11
  2:*     **                        11 12
  1:*     42
  0:*
IN H.D.> 5    10   15   20   25   30   35   40   45

 COLUMNS: 0  5  6  7 31 32 33 34 35 36
AVERAGES: 5  8  7  6 10 12 10 11 13 12
```

Figure 4: Example of part of program output: Scatter diagram of output Hamming distances against input Hamming distances. The number of occurrences is posted in each cell, with an asterisk indicating that there are more than 9.

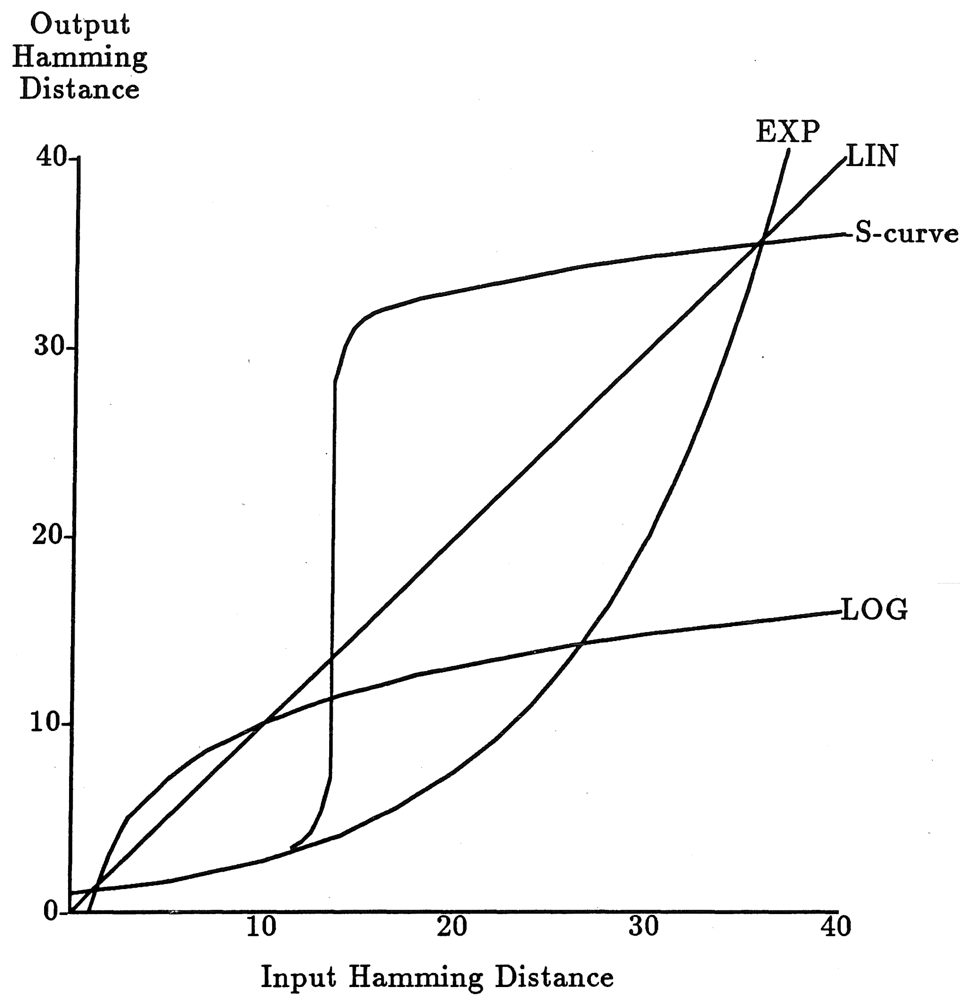

Figure 5: Some forms that Hamming Distance Graphs could show.

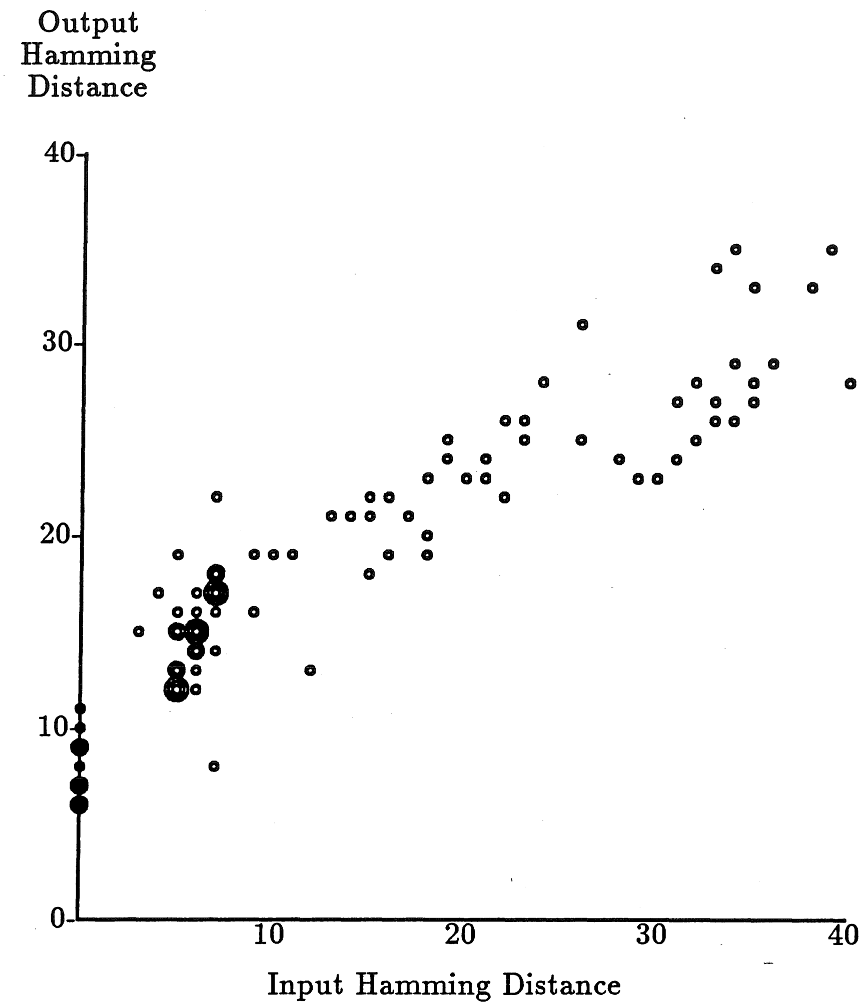

Figure 6: Linear relationship found between output distances and input distances for one set of neurobiological parameters. Concentric circles indicate multiple occurrences at that point.

5.1 Representational Flexibility

Any system that constructs unique internal representations is going to reflect trade-offs between conflicting performance goals. One trade-off is between *generalization power* and *discriminability.* Both capabilities are desirable, but a system that produces a single output for each input cannot satisfy varying goals depending on circumstances. A closely-related trade-off exists between the *representational capacity* of the system and its *robustness* (tolerance of noisy inputs).

A system capable of more flexible performance must be able to produce differing representations as requirements vary. Such a system must be capable of responding differently if it receives a signal that a robust/generalized representation is needed than if a *discriminate* signal is received, and intermediate responses are also desirable.

The pyriform cortex could attain this capability from its inhibitory network. The inhibitory cells receive inputs from other brain regions that greatly affect the cells responsiveness to L.O.T. and pyriform activity. This graded responsiveness increases or dampens the pyriform's output, and there may also be subtler qualitative effects that produce the desired performance trade-offs. Preliminary results are heavily dependent on the particular experimental protocol being simulated, but the system has displayed this ability under selected conditions.

5.2 Categorization

Another interesting property of a system that develops representations is categorization. In a system like the pyriform with many independent features for input and output, categorization may be defined in terms of the satisfaction of one or both of two conditions:

1. Many *very* similar inputs map to one output, as in the low range of the S-curve Hamming Distance function.
2. Many similar overlapping inputs map to a small overlapping set of outputs, and this output overlap constitutes the category. This con-

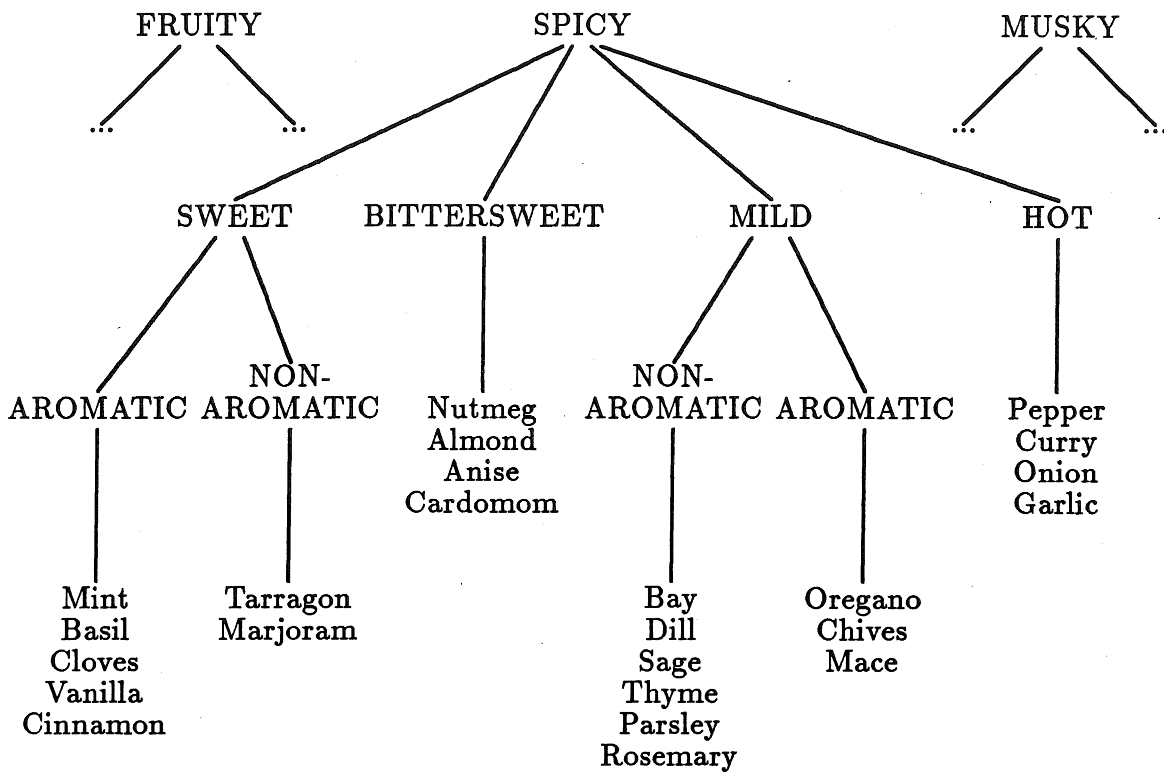

Figure 7: Example of Multiple Levels of Categories

dition is too specific to be measured by Hamming distance; instead, the category utility measure, mentioned earlier, can be used.

When these conditions are met at several levels, a hierarchy of two or more layers of categories will result (Figure 7). A few simulations have yielded such a multiple-level hierarchy, but it has not yet been tested over the full range of input presentation conditions.

Table 1: Object–Attribute–Value Representation Example

OBJECTS:	ATTRIBUTES:	VALUES:
Racketball	Size	Small
Softball	"	Medium
Basketball	"	Large
Plum	Color	Dark
Cantalope	"	Tan
Pumpkin	"	Orange
	Smell	Rubbery
	"	Leathery
	"	Fruity
	"	Pumpkin-y

6 Implications for AI Knowledge Representations

Most Artificial Intelligence (AI) systems maintain and operate on an internal representation of knowledge. A common knowledge representation formalism is the "object – attribute – value" encoding. Under this system, each describable object has only specified symbolic attributes, and each attribute has a pre-defined range of values. An example is pictured in Table 1. Allowing continuous real values for attributes is not common because of the increased computational demands in value-manipulating operations. In a few cases the systems are able to form simple compositions of the current attributes or values, but more typically, only the user-provided primitives are used.

Because the number of objects, attributes, and values must be limited in a tractable implementation, there is strong incentive not to "clutter" a system with items that play no direct role in the intended progress of the AI demonstration. This pre-selection limits the generalizability of the system's capabilities, however, since the system has had little opportunity to learn which items are salient.

The research reported in this paper suggests that some sensory areas of the brain may implement a clustering algorithm to avoid this limitation. The reported system "recognizes" common co-occuring patterns in the environment, and constructs an internal definition for an *attribute* as the set of values of arbitrary sensory primitives that co-occur. If such a clustering preprocessor were added to AI systems, the developer would no longer need to specify the set of values defining the attribute.

7 Conclusions

This paper has attempted to show how simulations of a system at the neural level, coupled with a theory of the function of the system, can provide evidence about the algorithm being used to perform the function.

Some early results of this effort were presented here. They provide some evidence that this approach might eventually yield a detailed understanding of the operation of both pyriform cortex and an artificial sensation-organizer and representation-learner. But much work remains before this goal can be accomplished.

References

[Carbonell et al. 83] Carbonell, J.G., Michalski, R.S., & Mitchell, T.M. (1983) An overview of machine learning. In Michalski, R.S., Carbonell, J.G., & Mitchell, T.M. (Eds.) *Machine Learning: An Artificial Intelligence Approach.* Palo Alto: Tioga.

[Fisher 87] Fisher, D. (1987) Knowledge acquisition via incremental conceptual clustering. *Machine Learning,* in press.

[Gluck & Corter 85] Gluck, M.A. and Corter, J.E. (1985) Information, uncertainty, and the utility of categories. *Proceedings of the Eighth Annual Conference of the Cognitive Science Society.* Hillsdale, N.J.: Lawrence Erlbaum & Associates.

[Hebb 49] Hebb, D.O. (1949) *The Organization of Behavior.* New York: Wiley.

[Kohonen et al. 81] Kohonen, T., Oja, E., and Lehtio, P. (1981) Storage and processing of information in distributed associative memory systems. In Hinton, G.E. and Anderson, J.A. (Eds.) *Parallel Models of Associative Memory.* Hillsdale, N.J.: Lawrence Erlbaum & Associates.

[Lynch 86] Lynch, G. (1986) *Synapses, Circuits, and the Beginnings of Memory.* Cambridge, Mass: MIT Press.

[Marr 82] Marr, D. (1982) *Vision.* New York: W.H. Freeman.

[Mervis & Rosch 81] Mervis, C. & Rosch, E. (1981) Categorization of natural objects. *Annual Review of Psychology,* 1981, pp. 89-115.

[Michalski & Stepp 83] Michalski, R.S. & Stepp, R.E. (1983) Learning from observation: conceptual clustering. In Michalski, R.S., Carbonell, J.G., & Mitchell, T.M. (Eds.) *Machine Learning: An Artificial Intelligence Approach.* Palo Alto: Tioga.